Analysis of California Wildfires From 1985-2015

Institution: Foothill College

Course: BUS 12 Data Analytics and Business Decisions

Student: Pramukh Bhushan

Table Of Contents

Project Summary

Domain Frame

California wildfires were analyzed in this project due to their severe impact on Californians in recent years. According to the *NY Times*, "10 of the largest fires occurred since 2000," including the biggest fire—the 2018 Mendocino Complex—and the sixth largest fire—the 2018 Carr fire—in state history (Pierre Louis & Schwartz). Wildfires have spurred into megafires, wreaking havoc across the state. For the economy, fires have destroyed "18,000 structures, 54% of which were homes," "[caused] 28 Billion in capital losses [in 2018]," and nearly "$150 billion in damages" in 2018 alone (Wibbenmeyer & McDarris). Not only are these unrestrictive fires catastrophic for humans, but they are also detrimental to the environment: wildfires "impact water quality and water supply" due to contamination with toxic chemicals, destroy old growth trees, and in 2020 alone, added, "112 million metric tons of carbon dioxide" (Wibbenmeyer & McDarris). California needs to find new solutions for combatting large fires and implement more preventive solutions. One preventative measure this project mentions is ecological forest restoration, a method that primes forests to survive extreme fire by taking out brush, unhealthy trees, and performing control burns (Let's Fight Fire with Fire).

Problem Statement/Hypothesis

This project's central question is whether "California wildfires are getting worse and what can we do to stop it." The relationships and visualizations I created also answer the sub-questions: "are wildfires being easier to put out," "what causes the deadliest fires," and "which agency deals with the most severe fires." Some potential uses for my findings are

conveying the insights found in safety warning brochures and understanding where California should increase funding regarding wildfires.

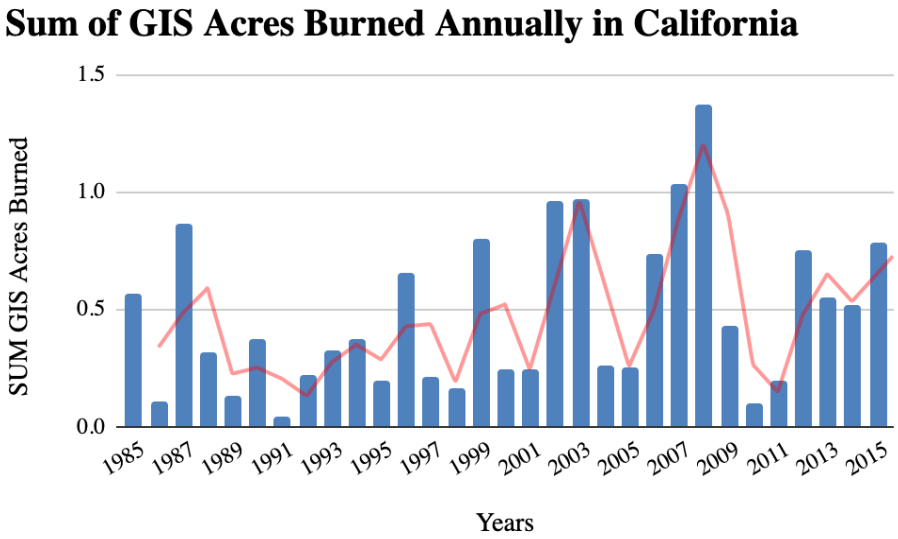
Data

The figures used in this project are a summary of the wildfires listed in Cal. Fire directory and can be found on the Cornell Virtual Website Workshop. However, my data analysis only analyzed information from 1985 to 2015. The elements analyzed were: year, agency, cause, GIS_ACRES, ALARM_DATE, CONT_DATE (containment date), and Fire length. The data set was cleaned by deleting data about other states, changing NULL values to the "14-Unknown/Unidentified" label, creating a formula to calculate fire length in days, and deleting rows with fires lasting 31,000 and 3,000 days. Before the data cleansing, there were 17 Columns and 3,897 rows; after, there were 18 columns and 3,871 rows.

Exploratory Data Analysis/Findings

Graph 1: Upward Trend for Acres Burned due to California Wildfires Overtime

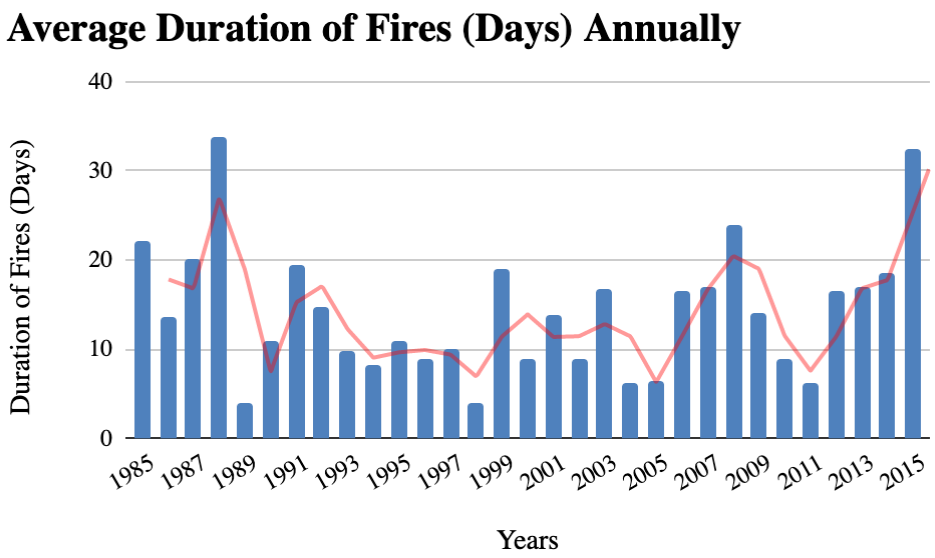
Recommendation: Create more buffers and warnings in areas with high fire zones



In graph 1, acres burned are generally increasing yearly, meaning there is an upward trend for the graph. The increase in acres burned by wildfires annually indicates that wildfires are indeed getting worse and may be due to climate change, increased population, or people building structures on the hills. A potential solution could involve creating more buffers like baseball or football fields so the fire could die out faster before reaching many buildings. Another insight is that there are significant spikes in the graph for 2003 and 2008, indicating that those years had many more acres burned than others. The large spikes could have been caused by megafires like the Cedar fire in 2003.

Graph 2: Decline In Average Duration of Fires Until 2012

Recommendation: Invest In New Technology and Ecological Forest Restoration

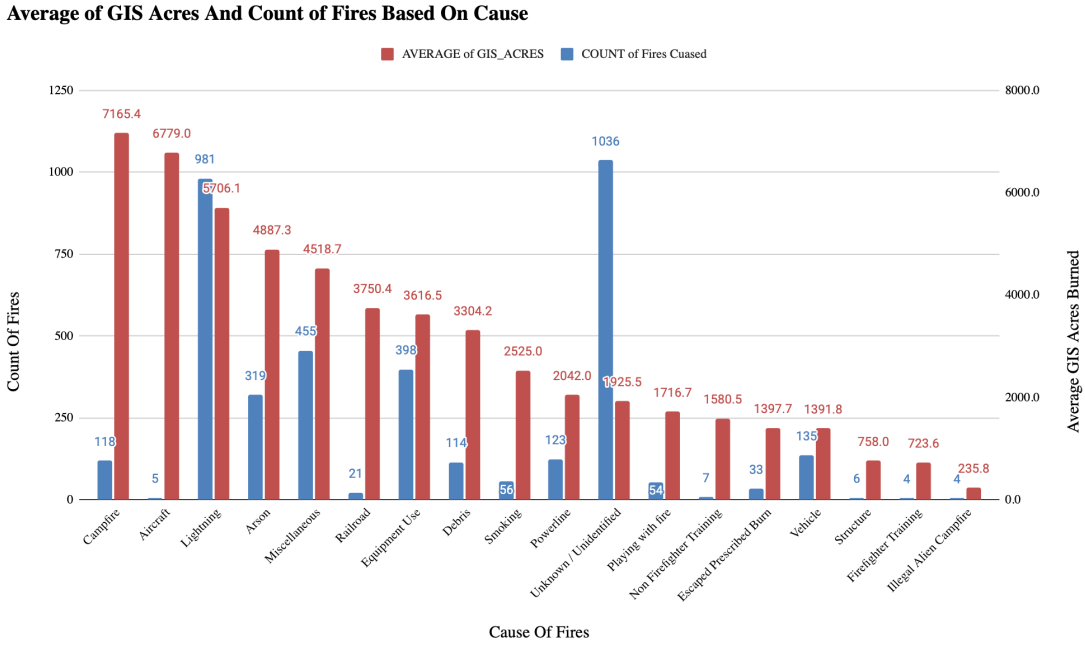


The initial downward trend in the graph means fires were contained much faster (high efficiency) prior to 2012, and the upward trend means fires after that year took longer to extinguish (less efficiency). Overall, this means that fires are getting worse as they are taking longer for

firefighters to stop. One factor for the increase in the duration of fires in 2012 could be the result of the Rush Fire that took place that year, the 10th largest fire that occurred in California. By investing in ecological forest restoration and research, mega-fires like the Rush Fire can be controlled in the future.

Graph 3: The Second Highest Cause of Fires is Lightning, and Campfires Cause the Most Damage

Recommendation: Restrict camp zones in high fire areas & increase forest restoration

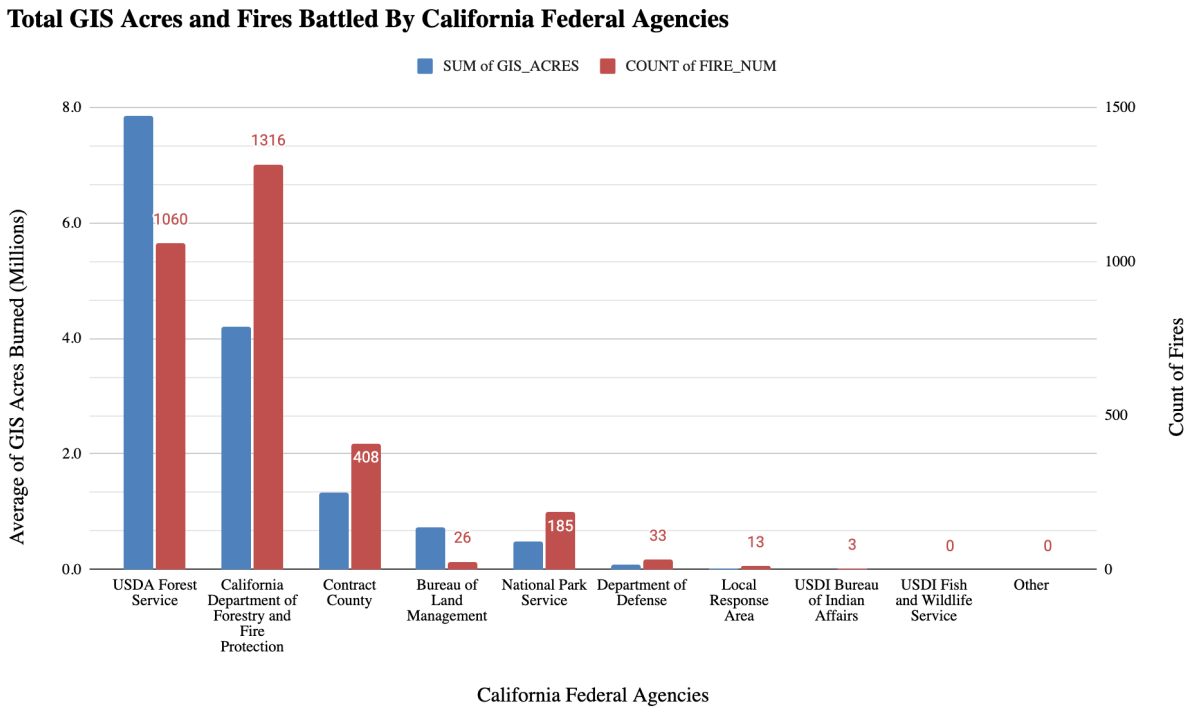


The second highest cause of fires is lightning, and campfires caused the most acres burned. One potential explanation is that campfires are built in forests, an environment where fire can spread more quickly. An increase in lightning could be due to climate change, but forest restoration can reduce the chance of sparks destroying the entire forest. Thus, restricting camp zones in high

forest areas and implementing forest restoration can reduce the severity of the leading causes of wildfires.

Graph 4: USDA Forest Fires battled fires that with the highest average acres burned, and the California Department of Forestry battled the most number of fires.

Recommendation: Provide more funding for USDA Forest Service and the California Department of Forestry



Looking at the bars, USDA Forest Fires battled fires that burned the highest average acres, and the California Department of Forestry battled the most number of fires. One potential

explanation is that they are the biggest agencies or cover the most “at risk” areas. By funding these agencies better, California can prevent more destruction caused by wildfires.

Reflection

SELECT Domain

Learning about California wildfires was relatively easy as it is a heavily discussed topic on the news. Some great resources I used were *L.A. Times* and the California Fire website. As a resident of California, this issue was close to my heart.

Formulate/Frame

The hypothesis was formulated based on essential questions I had about wildfires. However, I could not cover all the aspects of the questions, so exploring the variables in the data set gave me a better understanding of what could actually be analyzed. One aspect of the project that was frustrating was figuring out the right KPI for the question as I realized that years could be compared with the acres burned, the time it took to extinguish, or the number of fires that occurred that year.

Obtain

The data set was selected after examining databases from Kaggle, *data.com*, and the Cornell virtual workshop. Kaggle had quite a bit of data, but I felt it was incomplete because there were many missing/blank values. *Data.com* had a similar data set to Kaggle, so I believe it is from the same source. Finally, I stumbled on a data set from a credible university, Cornell, with all the columns I needed with minimal missing points. This data set was not collected by Cornell but was a summarized version of the excel sheet from the California Fires Database.

Scrub

I selected a range of data that covered 30 years (1985-2015) because the data set I found had 11,000+ rows. After finding a range of years, I discovered and deleted information about states that were not California. Next, I changed NULL values in the "Cause" column to 14-Unknown/Unidentified to better fit the data category. One lucky break I had was when I realized I could calculate the time it took to extinguish a fire by subtracting "CONT_DATE" and "ALARM_DATE." I also deleted illogical data like end dates listed before the start date and a row with a fire lasting 31,000 days and 3,00 days. One aspect of the process I found frustrating was that the pivot tables were not automatically sorted from largest to smallest.

Explore

Initially, I found insights quite quickly because I clearly outlined the questions, relationships, and KPIs I wanted to look at. However, halfway through writing the write-up, I realized some of my graphs did not make sense because I used "average" instead of "sum" on the pivot table. This was frustrating because I did not understand my analysis for some time and had to ask for second opinions from my peers.

Present

Presenting my data was pretty straightforward because my data fit the bar-graph format quite well. However, because my data was not a scatter plot, looking at linear trends was not my first option, making me look at specific bars in the graphs to draw insights.

Reflection: Key Learnings

Domain

In the future, I should do more background research before finding a domain so the analysis process can go faster.

Question or hypothesis

For the hypothesis and questions, I should try to make a more specific question and include two types of domains (i.e. Elon musk tweets and dogecoin), so I can draw unique connections between them later.

Obtaining Data

I think I got lucky with my dataset, but in the future, I should look at at least 3-6 different domains to see if the columns align with my question rather than forming questions based on the available data. For the cleaning process, I should define what is considered extraneous data before the cleaning process.

Present

Next time I explore the data, I want to use different types of graphs that are optimized with the analyzed data from a specific range of values.

Conclusion:

Overall, I enjoyed this project and learned much about data analysis and California wildfires. However, I feel there were some limitations with my research because the number of fires each year in my data sets did not match the information I found on Google. Next time I will analyze databases with more rows and make sure to verify it with information online.

References

California: Let's Stop Making Wildfire History. (2022, July 15). The Nature Conservancy. https://www.nature.org/en-us/about-us/where-we-work/united-states/california/stories-in-california/californias-wildfire-future/

Historical California Wildfire Data. Cornell University. https://cvw.cac.cornell.edu/PyDataSci1/wildfires

Matthew Wibbenmeyer, M. W., & McDarris, A. M. (2021, July 31). *Wildfires in the United States 101: Context and Consequences*. Resources for the Future. https://www.rff.org/publications/explainers/wildfires-in-the-united-states-101-context-and-consequences/#:%7E:text=In%202020%2C%20wildfires%20destroyed%20almost,to%20both%20homes%20and%20businesses

Pierre-Louis, K., & Schwartz, J. (2021, July 16). *California Wildfires Explained: Why Does The State Have So Many?* The New York Times. https://www.nytimes.com/article/why-does-california-have-wildfires.html